Weekly Tip for Oct. 20, 2020

Check your statistical assumptions!

Regression analyses are one of the main steps (aside from data cleaning, preparation, and descriptive analyses) in any analytic plan, regardless of plan complexity. Therefore, it is worth acknowledging that the choice and implementation of the wrong type of regression model, or the violation of its assumptions, can have detrimental effects to the results and future directions of any analysis. Considering this, it is important to understand the assumptions of these models and be aware of the processes that can be utilized to test whether these assumptions are being violated.

Keep in mind, each model has its own set of assumptions that must be met! For example:

- Common Assumptions of Parametric Tests: normality, homogeneity of variance, homogeneity of variance-covariance matrices, linear relationships, absence of multicollinearity, absence of autocorrelation, and randomization
- Logistic Regression Assumptions: dependent variable structure, observation independence, absence of multicollinearity, linearity of independent variables and log odds, and large sample size
- Linear Regression Assumptions: linearity, multivariate normality, absence of multicollinearity and auto-correlation, homoscedasticity, and measurement level

If you find that a model assumption has been violated, have no fear! Most assumptions can be corrected with minimal impact on the interpretability of your results. Corrections can range from minor value transformations and the exclusion of noisy variables or observations to the consideration of a more appropriate model type.

Don’t forget to test for assumptions! If you run a model with data that does not match its theoretical assumptions, your results may cause more harm than running no model at all!

This week’s tip was contributed by Deanna Schreiber-Gregory. Deanna is a government contractor and independent consultant who specializes in statistics, research methods, and data management.

Weekly Tip for Nov. 10, 2020

Assessing the performance of a statistical model

In the submission/presentation phase of any research or analytics project, it is reasonable to expect the reception of many types of questions aimed at clarifying the reliability and accuracy of the project's results. One of the most common questions to expect would be: "So the model provides a feasible answer to the question, but does it provide the best answer?" One way to answer this question with utmost confidence is to provide a variety of model fit analyses designed to support the conclusion of your final model. Two very common ways to explore model performance is to run tests for goodness-of-fit and predictive power.

- Goodness-of-fit: Goodness-of-fit measures are formal tests of the null hypothesis that the fitted model is correct. These measures output a p-value which is used to decide whether or not the indicated model is a good fit. P-values are numbers between 0 and 1 with higher values indicating a better fit. Contrary to the traditional view of p-values, where one would specify a target α level (such as .05) and accept a model with a p-value below this value, goodness-of-fit test p-values that land below the specified alpha level would indicate that a model is not acceptable.

  Types: AIC, BIC, -2LogL, Stukel's test, Information Matrix Test, Unweight Sum of Squares, Standardized Pearson Test, Hosmer-Lemeshow, calculations of deviance, Pearson Chi-Square

- Predictive Power: Measures of predictive power typically have values that fall between 0 and 1, with 0 indicating a complete lack of predictive power and 1 indicating a perfect predictive relationship. As a general rule, the higher the value, the better, but other than that there are rarely any fixed cut-off values that differentiate whether a model is acceptable or not.

  Types: R-Square (Cox-Snell, Tjur), ROC, rank-order correlations (Somers' D or Gini coefficient, Goodman-Kruskal Gamma, and Kendall's Tau-a)

On a given model, you do not need to run all of these tests! However, it is always a good idea to explore the underlying theoretical basis of the model you are running and match it to the appropriate performance test. Remember, presenting a weak model without an assessment of its limitations may cause more harm than presenting no model at all!

This week's tip was contributed by Deanna Schreiber-Gregory. Deanna is a government contractor and independent consultant who specializes in statistics, research methods, and data management.

Weekly Tip for Jan. 19, 2021

Use SAS Survey Procedures for Complex Survey Analysis

It is common practice when learning statistics for students to distribute a simple survey of mundane content (What is your favorite color? What is your undergraduate major?) to collect data. This data is then compiled into a small data set and used for simple statistical exploration. Considering this, it is no wonder that we tend to assume that any statistical procedure can be applied to survey data. Though there is some truth to this frame of mind, we must acknowledge that the method of collecting survey data violates a number of our favorite procedure's most restrictive assumptions. As explained in an earlier Tuesday Tip, assumption violations effectively destroy a model's credibility if they are not addressed.

Since survey data is so widely used, SAS sought to address the common issues with survey analysis by developing a series of seven survey specific analytic procedures:

- PROC SURVEYSELECT: Used to select a sample from a data set.
- PROC SURVEYIMPUTE: Used to do single imputations on a survey data set.

- PROC SURVEYMEANS: Used to obtain weighted descriptive statistics for continuous variables and produce accompanying graphics.
- PROC SURVEYFREQ: Used to run weighted one-way and multi-way cross-tabulations and produce accompanying graphics.
- PROC SURVEYREGRESS: Used to run weighted OLS regressions.
- PROC SURVEYLOGISTIC: Used to run weighted logistic, ordinal, multinomial and probit regressions.
- PROC SURVEYPHREG: Used to run weighted proportional hazards regression.

It is important to make sure that you are using the appropriate procedure for the data you have. If you are working with survey data, please make sure to take a look at these procedures. They have built in options that allow you to manipulate your model to reduce the impact of common assumption violations.

Happy analyzing!

This week's tip was contributed by Deanna Schreiber-Gregory. Deanna is a government contractor and independent consultant who specializes in statistics, research methods, and data management.

Weekly Tip for Feb. 9, 2021

Consider Latent Structure Analyses to Find Hidden Variables

When being introduced to statistics, students are usually presented with a group of variables to analyze. One or two variables are identified as being the subjects or dependent variables of the analysis, while the others are identified as the independent variables or contributing factors. A specific statistical model is then run by following the directions outlined in the particular problem and a result is produced. This process enables students to develop a grasp on how variables in the world interact and affect each other. However, as most of us know, there is much more going on in the interactions of objects and people in daily life than can be either simply observed or explained by any single identified variable. Therefore, how can we obtain a grasp on those more subtle, unobserved aspects of cause/effect and interaction that are otherwise too difficult to measure?

One way to look at "unobserved" variables is through latent variable modeling. Through this type of modeling, a statistician is able to view the impact of variables that are not able to be directly observed during the course of a study. Latent variables are included in many different kinds of regression models and are more formally referred to as "systematic unmeasured variables" or factors.

Latent structure analyses are the set of theoretical analytic procedures developed to find, measure, and define potential latent variables in a data set. Considering the positive effect on model quality that can result from the inclusion of latent variables, their nature and application certainly warrants further exploration. However, not all latent variables can be calculated the same. Whether the variable is created from categorical data, continuous data, or data represented across time, different latent variable analyses must be taken into consideration.

Here are some examples of common latent structure analysis procedures:

1. For data that takes on a categorical nature, a ***latent class analysis*** would be used to help identify latent class variables with this type of format.

2. For data that it represented in a continuous format, a ***latent profile analysis*** would be the appropriate application.
3. For data that represents points across time, a ***latent transition analysis*** or ***latent trajectory analysis*** are necessary to explore the latent variables that appear over time.

Structure equation modeling procedures, such as SAS's PROC CALIS, are uniquely equipped to handle latent structure analysis demands. For more fine-tuned procedures, several statisticians have produced and maintained packages and add-ons for a variety of analytic platforms to help fellow researchers get to the bottom of their complex data sets.

I strongly encourage you to check out these procedures, as their ability to uncover hidden interactions within your data set is a powerful addition to any project!

This week's tip was contributed by Deanna Schreiber-Gregory. Deanna is a government contractor and independent consultant who specializes in statistics, research methods, and data management.

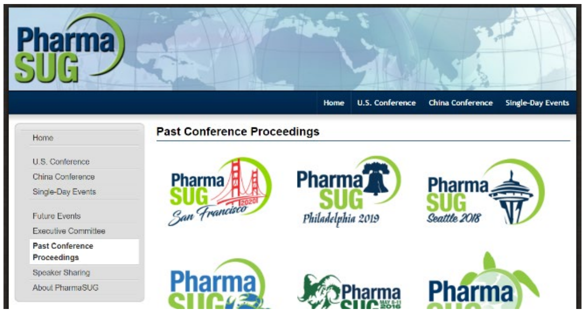
Weekly Tip for Feb. 16, 2021

Where to look for help with life science statistics, and what to do when you find it!

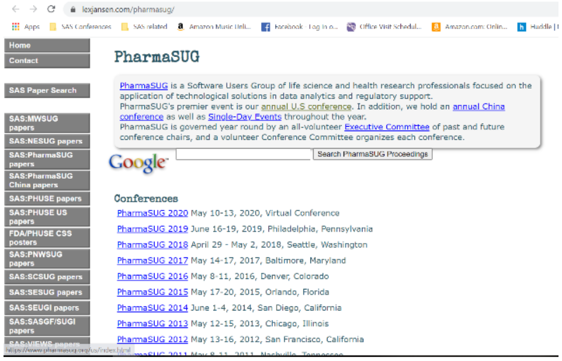
This tip is one of the most useful tips you will encounter! The technical information was not written by the provider, but it was contributed, nonetheless. When Lauren Rackley went looking for some information on Kaplan-Meier survival plots for chart time to recovery, she found this PharmaSUG paper by Jeffrey Meyers: https://www.lexjansen.com/pharmasug/2014/BB/PharmaSUG-2014-BB13.pdf. The paper (and macro) allowed her to accomplish her work task far more easily than reinventing the wheel, allowing her to take advantage of GTL to show side-by-side comparisons of survival analyses.

Jeffrey has developed, and shared with PharmaSUG readers, a macro NEWSURV which leverages Graph Template Language to provide journal-ready graphical representation of median time-to-event, number of patients, and time-point event-free rate estimates inside a Kaplan-Meier plot. Additionally, his macro allows for side-by-side comparisons of plots, using the Lattice layout in GTL. His paper is well worth the read.

Where do you go to find gems like this paper? All it takes is a few keystrokes into a search engine such as Google, but it's much better to start your search in a more targeted manner. If you are reading this tip, you have already found your way to the PharmaSUG website. PharmaSUG maintains proceedings for a number of PharmaSUG conferences online.



Go to PharmaSUG's past conference proceedings (https://www.pharmasug.org/proceedings.html), and click and peruse papers in conference sections you are interested in. If you don't find what you are looking for in the last 10 years or so worth of proceedings, one of PharmaSUG's own, Lex Jansen, has also created a fantastic library of virtually all past PharmaSUG and other conferences as well at https://www.lexjansen.com/pharmasug/.



lexjansen.com/pharmasug/

Apps | SAS Conferences | SAS related | Amazon Music Unli... | Facebook - Log In o... | Office Visit Schedul... | Amazon.com: Onlin... | Huddle |

Home
Contact

SAS Paper Search

SAS:MWSUG papers
SAS:NESUG papers
SAS:PharmaSUG papers
SAS:PharmaSUG China papers
SAS:PHUSE papers
SAS:PHUSE US papers
FDA/PHUSE CSS posters
SAS:PNWSUG papers
SAS:SCSUG papers
SAS:SESUG papers
SAS:SEUGI papers
SAS:SASGF/SUGI papers
SAS:VIEWS papers

https://www.pharmasug.org/us/index.html

PharmaSUG

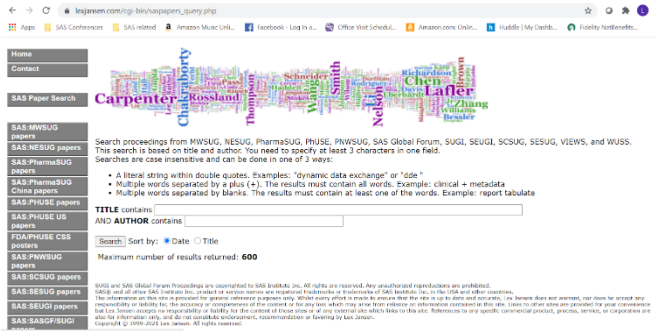
PharmaSUG is a Software Users Group of life science and health research professionals focused on the application of technological solutions in data analytics and regulatory support.
PharmaSUG's premier event is our annual U.S conference. In addition, we hold an annual China conference as well as Single-Day Events throughout the year.
PharmaSUG is governed year round by an all-volunteer Executive Committee of past and future conference chairs, and a volunteer Conference Committee organizes each conference.

Google [ ] Search PharmaSUG Proceedings

Conferences

PharmaSUG 2020 May 10-13, 2020, Virtual Conference
PharmaSUG 2019 June 16-19, 2019, Philadelphia, Pennsylvania
PharmaSUG 2018 April 29 - May 2, 2018, Seattle, Washington
PharmaSUG 2017 May 14-17, 2017, Baltimore, Maryland
PharmaSUG 2016 May 8-11, 2016, Denver, Colorado
PharmaSUG 2015 May 17-20, 2015, Orlando, Florida
PharmaSUG 2014 June 1-4, 2014, San Diego, California
PharmaSUG 2013 May 12-15, 2013, Chicago, Illinois
PharmaSUG 2012 May 13-16, 2012, San Francisco, California
PharmaSUG 2011 May 8-11, 2011, Nashville, Tennessee

If you are looking for help with SAS programming, you can go one step further and use the SAS paper search functionality. This allows you to search and download results in JSON or XML.



lexjansen.com/cgi-bin/saspapers_query.php

Apps | SAS Conferences | SAS related | Amazon Music Unli... | Facebook - Log In o... | Office Visit Schedul... | Amazon.com: Onlin... | Huddle | My Dashb... | Fidelity NetBenefits...

Home
Contact

SAS Paper Search

SAS:MWSUG papers
SAS:NESUG papers
SAS:PharmaSUG papers
SAS:PharmaSUG China papers
SAS:PHUSE papers
SAS:PHUSE US papers
FDA/PHUSE CSS posters
SAS:PNWSUG papers
SAS:SCSUG papers
SAS:SESUG papers
SAS:SEUGI papers
SAS:SASGF/SUGI papers

Search proceedings from MWSUG, NESUG, PharmaSUG, PhUSE, PNWSUG, SAS Global Forum, SUGI, SEUGI, SCSUG, SESUG, VIEWS, and WUSS.
This search is based on title and author. You need to specify at least 3 characters in one field.
Searches are case insensitive and can be done in one of 3 ways:

- A literal string within double quotes. Examples: "dynamic data exchange" or "dde "
- Multiple words separated by a plus (+). The results must contain all words. Example: clinical + metadata
- Multiple words separated by blanks. The results must contain at least one of the words. Example: report tabulate

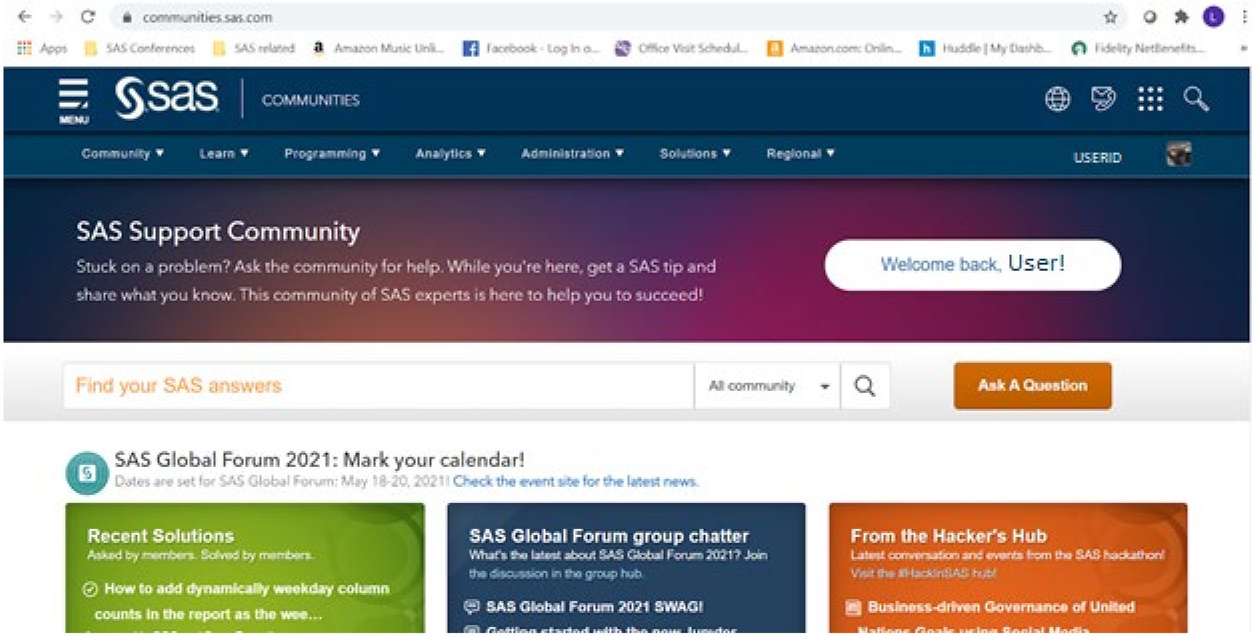
TITLE contains [ ]
AND **AUTHOR** contains [ ]

Search Sort by: ● Date ○ Title

Maximum number of results returned: **600**

SUGI and SAS Global Forum Proceedings are copyrighted to SAS Institute Inc. All rights are reserved. Any unauthorized reproductions are prohibited.
SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.
The information on this site is provided for general reference purposes only. Whilst every effort is made to ensure that the site is up to date and accurate, Lex Jansen does not warrant, nor does he accept any responsibility or liability for, the accuracy or completeness of the content or for any loss which may arise from reliance on information contained in this site. Links to other sites are provided for your convenience but Lex Jansen accepts no responsibility or liability for the content of those sites or of any external site which links to this site. References to any specific commercial product, process, service, or corporation are also for information only, and do not constitute endorsement, recommendation or favoring by Lex Jansen.
Copyright © 1999-2021 Lex Jansen. All rights reserved.

SAS also provides proceedings from SAS conferences in the SAS Support Community website (communities.sas.com) which allows you to search topics and to engage in Q and A with experts all over the world. Many of the tips on Communities.sas.com also have examples in GitHub.



communities.sas.com

Apps | SAS Conferences | SAS related | Amazon Music Unli... | Facebook - Log In o... | Office Visit Schedul... | Amazon.com: Onlin... | Huddle | My Dashb... | Fidelity NetBenefits...

MENU

Community ▾ | Learn ▾ | Programming ▾ | Analytics ▾ | Administration ▾ | Solutions ▾ | Regional ▾ | USERID

SAS Support Community

Stuck on a problem? Ask the community for help. While you're here, get a SAS tip and share what you know. This community of SAS experts is here to help you to succeed!

Welcome back, User!

Find your SAS answers | All community ▾ | Ask A Question

SAS Global Forum 2021: Mark your calendar!
Dates are set for SAS Global Forum: May 18-20, 2021! Check the event site for the latest news.

Recent Solutions
Asked by members. Solved by members.
How to add dynamically weekday column counts in the report as the wee...

SAS Global Forum group chatter
What's the latest about SAS Global Forum 2021? Join the discussion in the group hub.
SAS Global Forum 2021 SWAG!
Getting started with the new Invoder

From the Hacker's Hub
Latest conversation and events from the SAS hackathon! Visit the #HackinSAS hub!
Business-driven Governance of United Nations Goals using Social Media

Additionally, the various social media platforms (LinkedIn, Twitter, Facebook, Instagram, Google and others) provide good search engines and the ability to link up with subject matter experts.

There are also special interest groups and list-servs that can provide help, such as the venerable SAS-L and newer groups such as the SASensei group on LinkedIn.

Lauren notes that Jeffrey Meyer, author of the paper she recommended as a great tip, was very helpful when she contacted him with questions. Most paper authors take pride in their work and are happy to extend a helping hand and exchange ideas. We highly recommend searching through the PharmaSUG conference proceedings (https://www.pharmasug.org/proceedings.html) as a good starting point when you have questions! The PharmaSUG social media team also sponsors “Friday Faves” which profile favorite PharmaSUG papers so if Lauren’s tip has inspired you, stay tuned to both Tuesday Tips and Friday Faves! #PharmaSUGTuesTip #PharmaSUGFriFave

Lauren Rackley is a Statistical Programmer in the Data Science & Analytics Group at DLH Corporation. In this role, she creates TFLs for an NIH Phase II clinical trial.